Spectral and temporal properties of Estonian palatalization

Anton Malmi [ID]

University of Tartu, Estonia
anton.malmi@ut.ee

Pärtel Lippus [ID]

University of Tartu, Estonia
partel.lippus@ut.ee

Einar Meister [ID]

Tallinn University of Technology, Estonia
einar.meister@taltech.ee

The aim of this paper is to describe and compare the spectral and temporal properties of Estonian palatalized and non-palatalized consonants /l n s t/ and the vowels that precede them. Acoustic recordings of 43 native Estonian subjects producing word pairs where palatalization differentiated meaning were analyzed in this study. We offer a new perspective on how palatalization is realized by employing a dynamic analysis of the formant trajectories in order to objectively quantify how far the scope of palatalization reaches in the preceding vowel. Results showed that, as in other languages, the most persisting correlate of Estonian palatalization is the rise in F2 in the preceding vowel, where the values are almost always higher already from the beginning. F2 values are higher for /l/ and lower for /n/. The COG (center of gravity) of /s/ is lower in the beginning but rises in the middle. The COG of the burst of /t/ is lower in the onset. As expected, the duration of the vowels preceding consonants is longer, presumably because of the palatalization gesture, but the duration of the consonants does not show a systematic pattern with palatalization.

1 Introduction

This paper aims to present a study on the spectral and temporal properties of palatalization in Estonian consonants and in the vowels that precede them. Palatalization is a complex phonological process common in world languages. It can be found in a variety of languages such as English, Russian, Japanese, French, Mandarin Chinese, Greek, and elsewhere, including Estonian (Bateman 2011). The term palatalization can cover various underlying phonological processes like coronalization or fronting (Hume 1994), velar softening (Chomsky & Halle 1968), or spirantization (Bhat 1978), where a coarticulatory assimilation of a consonant and a neighboring high front vowel or a glide /j/ occurs (Kochetov 2011). The trigger, the target, and the outcome of palatalization can vary cross-linguistically. For example, it is common for high and front vowels or glide /j/ to trigger palatalization, but it has been shown that, for

example, in Tohono O'Odham, Coatzospan Mixtec, Maori, and Sentani languages that high back and central vowels can trigger palatalization as well (Bateman 2007).

Palatalization can be divided into full and secondary types by looking at the change in the target consonant since the trigger generally does not determine the type of palatalization. With full palatalization, the consonant changes its place and manner of articulation while moving to the palatal region of the mouth. With secondary palatalization, the consonant's main place of articulation remains the same, but the tongue raising during the articulation of the following segment gives the consonant a secondary place of articulation on the hard palate. Secondary articulations are common, and the secondary raising of the tongue or retracting the tongue root can make the consonant palatalized, velarized, uvularized, pharyngalized, or laryngalized (for an overview, see Catford 1977: 188–196). This article will focus on the secondary palatalization of four Estonian coronal consonants. Labial and dorsal consonants do not palatalize in Estonian. It is not uncommon that only coronal consonants palatalize in a language (Bateman 2007).

The typological distribution of the palatalization targets varies. They can involve coronal and dorsal consonants independently or together, and labial consonants can be targeted only when coronals and dorsals are targeted in a language (Kochetov 2011). The position of the trigger can precede or follow the target consonant, but it has been found that prepalatalization, which also exists in Estonian, is more common (Bhat 1978, Bateman 2007, Kochetov 2011). Regardless of the process described, the result is that the consonant's place of articulation is changed or the consonant becomes palatalized because of the nearby vowel or glide; this results in the consonant becoming more similar to the palatalization trigger. In a more recent study of Brazilian Portuguese, Battisti & Hermans (2020) argue that palatalization is structurally motivated, and the process affects consonants that are similar to the trigger vocoids because high vocoids have a high degree of consonantality.

We chose Estonian, a Balto-Finnic language of the Uralic language branch, for our study because Estonian has contrastive secondary palatalization in only four coronal consonants /l n s t/ and it presents a comparable binary dataset of minimal pairs in which palatalization differentiates meaning. The high variability of the possible targets and triggers and their outcomes motivated us to pursue the topic and to assess whether palatalization in Estonian follows the same patterns as it does in other languages. This article builds upon the knowledge and uses data from the same test subjects and recordings used in previous articles, namely a preliminary study preceding this paper by Malmi (2019) and an articulatory study by Malmi & Lippus (2019).

On the basis of palatalization, two series of alveolar phonemes can be defined in Estonian: non-palatalized /l n s t/ and palatalized /lʲ nʲ sʲ tʲ/ (Asu & Teras 2009). Estonian contrastive palatalization is realized in monosyllabic words after the first vowel and in disyllabic words on the boundary of the first stressed and following unstressed syllable (Kask 1972: 118; Teras & Pajusalu 2014). This results in an inventory of word pairs like *tulp* 'column' [tulpː] ∼ *tulp* 'tulip' [tulʲpː] or *kott* 'worn out shoe' [kotː] ∼ *kott* 'bag' [kotʲː]. These kinds of phonologically contrastive pairs were historically introduced into Estonian through the apocope of word final /i/ (Kask 1972: 118). Palatalization in Estonian is regressive, which means that the word final /i/ affected the place of articulation of previous alveolar consonant or all alveolar consonants in a cluster. Palatalization of a consonant was marked in orthography up until the beginning of the 20th century but got lost after the written language standard changed. This resulted in words with identical orthographic forms but different pronunciations. Some other pairs might not have a counterpart in the same grammatical case or the same quantity degree, but the contrast between the segmental sequences still remains, for example, *laste* 'child, PL.GEN' [lɑste] ∼ *laste* 'cargo, PL.PART' [lɑsʲːte]. The existence of such pairs allows us to compare and study the spectral and temporal properties of palatalization.

Phonetically, palatalization in Estonian is a result of regressive assimilation of neighboring segments in which the trigger for palatalization immediately follows the consonant target. However, in spontaneous speech, examples of progressive palatalization can be found as well,

e.g. when an /i/ in the second part of a diphthong (e.g. *kuidas* 'how' [kutʲɑs]) gets reduced or assimilated, the consonant is still articulated with palatalization (Teras & Pajusalu 2014). Non-contrastive palatalization in Estonian is, in many cases, optional and gradually variable, as the degree of palatalization can vary regionally or idiomatically (Teras & Pajusalu 2014, Põld 2016, Piits & Kalvik 2019). For this reason, in this study, only phonologically contrastive palatalization will be analyzed.

Acoustically, palatalization can be defined by the change in the quality of the vowel which precedes the palatalized consonant. A number of studies of different languages have found that the increase of the F2 value in the vowel preceding the palatalized consonant is the most persistent correlate of palatalization, but researchers have also observed that there is a decrease in F1 and an increase in F3 as well (Ní Chiosáin & Padgett 2012 for Connemara Irish; Lehiste 1965, Liiv 1965a, b, Vihman 1967, Teras & Pajusalu 2014 for Estonian; Derkach, Fant & De Serpa-Leitao 1970, Purcell 1979, Howie 2001, Kochetov 2002 for Russian; Kim 2012 for Korean; Ćavar 2004 for Polish). Vowels that have a higher F2 go through a smaller change in F2 than vowels with a lower F2 (Lehiste 1965). Similar patterns can be expected to occur in other languages that exhibit secondary palatalization because articulatorily, the process involves raising the body of the tongue to the hard palate. The transition from the vowel to the consonant is said to have an [i]-like quality compared to the non-palatalized context. The majority of those previous studies of palatalization have focused on describing the transition from the vowel to the consonant, or they have described the quality of the vowel and the consonant from a single or few time points.

Along with the changes in the formants of the preceding vowel, the effect of palatalization can be seen in the quality of the consonant itself as well. The center of gravity (COG; the average of the frequencies in a selected time window of the spectrum) of palatalized fricative consonants is lower for /s/ (Padgett & Żygis 2003, Hamann & Heriberto 2007, Kochetov & Radisic 2009, Malmi 2019). Labial and alveolar palatalized stops have higher frequencies in the spectrum of the burst (Shupljakov, Fant & De Serpa-Leitao 1968, Kochetov 2002, Kavitskaya 2006, Botinis, Chaida & Magoula 2011, Ní Chiosáin & Padgett 2012). /lʲ/ is described as having a higher F2 and F3 and lower F1 values compared to non-palatalized /l/ (Lehiste 1965, Vihman 1967, Remmel & Eek 1971, Eek 1972) and a higher COG (Malmi 2019). Researchers have had difficulties findings differences in F1 and F2 of non-palatalized and palatalized /n/ (Lehiste 1965, Iskarous & Kavitskaya 2018, Nance & Kirkham 2020). Because of the high dampening effects of the soft nasal cavity and the anti-formant that forms in the oral cavity, which overlaps with F2, nasal consonants are characterized by low intensity and wide bandwidths (Fant 1960, Derkach et al. 1970, Eek 1972, Tabain et al. 2016). Lehiste (1965) found that the F3 values of /n/ showed a tendency to be lower with palatalization, and Malmi (2019) found that the COG of /n/ was lower with palatalization. Researchers have stated that palatalization tends to only affect the beginning of Estonian consonants (Lehiste 1965; Liiv 1965a, b; Eek 1972, 1973). It is not known whether palatalization also affects the spectral properties of Estonian consonants in the midpoint. In addition to spectral properties, palatalization affects the temporal structure of the phonemes.

Research has shown that the duration of the vowels preceding palatalized consonants and the duration of /s/ and the burst of /t/ show a tendency to be lengthened (Liiv 1965a, Bolla 1981, Zsiga 2000, Ćavar 2004, Kavitskaya 2006, Kochetov 2006, Ordin 2010, Teras & Pajusalu 2014, Stoll, Harrington & Hoole 2015). This lengthening presumably comes from tongue raising and is also used to enhance the perception of palatalization (Ordin 2010). We had trouble with reproducing these findings in our articulatory study (Malmi & Lippus 2019), as the effect of vowel lengthening did not show a clear pattern. This topic will be revisited in the current study.

Speech is a dynamic process in which phonemes do not occur in isolation; they are coarticulated with nearby segments, and to some extent, they become similar. The term coarticulation describes an unconscious process where the movements of different articulators and targets overlap and interact with one another. Palatalization arises from coarticulation

of adjacent or nearby segments. Coarticulation can affect the quality of individual segments (Lindblom 1963a, b; Öhman 1966; Amerman, Daniloff & Moll 1970; Hammarberg 1976; Lisker 1978; Farnetani & Recasens 2013; Malmi 2014, 2016) but also the whole syllable (Xu 2020) or even extend beyond a syllable boundary (Öhman 1966). Coarticulation arises from the need for minimal articulatory effort in conveying information. While palatalizing consonants, the raising of the anteo-dorsum towards the hard palate already begins in the first part of the vowel and continues until the end of a palatalized consonant (Malmi & Lippus 2019).

We offer a new perspective into the realization of palatalization and suggest that the transition takes place throughout the whole vocalic segment rather than having a stable vowel target section followed by a transition as previously described. The current paper supplements the results from our articulatory study and provides a wider view on the realization of palatalization in Estonian and in other languages with secondary palatalization. Based on previous findings, the research questions and hypotheses of this study are as follows:

(i) Firstly, we want to observe possible acoustic cues that describe palatalization in Estonian vowels. We hypothesize that the vowels preceding palatalized consonants will have lower F1, higher F2, and F3 values. We also want to specify the properties of the so-called [i]-like the transition from the vowel to the palatalized consonant and to find a way to formally determine the beginning of the transition in order to compare it in palatalized and non-palatalized contexts. We hypothesize that, because of coarticulation, it might be difficult to find a clear transition following a stable 'target' segment, and we expect that the formants are continuously moving throughout the whole vowel.

(ii) Secondly, we want to observe the possible acoustic cues that describe palatalization in Estonian consonants. We hypothesize that with /l/, the F2 and F3 values will be higher, and F1 will be lower with palatalization. F2 and F3 values of /n/ will be lower with palatalization. The COG of /s/ will be lower, and the burst of /t/ will be higher with palatalization. We also want to quantify the scope of palatalization. We hypothesize that the effect of palatalization reaches further than just the beginning of the consonant, as has been previously stated. This hypothesis is also based on our previous articulatory study (Malmi & Lippus 2019), in which the tongue was higher with palatalization and which stayed higher throughout the consonant.

(iii) Thirdly, we want to know whether the vowel is lengthened by the palatalization of the consonant and whether the consonant is lengthened as well. In alignment with what previous research has shown, we hypothesize that the duration of the vowel, the duration of /s/, and the burst of /t/ are longer for palatalized than for non-palatalized consonants. We expect then that the durations of /l/ and /n/ will be longer too.

2 Materials and method

2.1 Materials used in the study

To study the spectral and temporal properties of palatalization, an acoustic experiment was conducted at the University of Tartu and Tallinn University of Technology with 43 native Standard Estonian speaking participants (20 male, 23 female), aged 20–78 years (median 29 years).

In this study, 18 monosyllabic word pairs with phonemic contrast in palatalization (i.e. a total of 36 words) were analyzed (Table 1). The test words were embedded in written carrier sentences and presented to the participants in a randomized order using the software SpeechRecorder (Draxler & Jänsch 2004). The dataset also included 55 sentences that were not analyzed in this study.

The test words were embedded in the middle of the carrier sentence followed by a comma and a word starting with *mi*-, e.g. *Kevadel sündis väike **tall*** [tɑlː], *mitte vasikas* 'In the spring

Table 1 Test word pairs from the experiment; for these, palatalization differentiates meaning. All nouns in the table are SG.NOM if not listed otherwise.

	Non-palatalized			Palatalized		
/l/	Mall	[mɑlː]	(name)	mall	[mɑlʲː]	protractor
	tal	[tɑlː]	3SG.GEN (pers. pron.)	tall	[tɑlʲː]	stable
	hall	[hɑlː]	frost	hall	[hɑlʲː]	gray
	mul	[mulː]	1SG.GEN (pers. pron.)	mull	[mulʲː]	bubble
	palk	[pɑlkː]	wage	palk	[pɑlʲkː]	log
	salv	[sɑlːv]	bin	salv	[sɑlʲːv]	ointment
	tulp	[tulpː]	column	tulp	[tulʲpː]	tulip
	sulg	[sulːg]	bracket	sulg	[sulʲːg]	feather
/n/	kann	[kɑnː]	jug	kann	[kɑnʲː]	toy
	on	[onː]	be-3SG.PRS	onn	[onʲː]	hut
/s/	kas	[kɑsː]	whether	kass	[kɑsʲː]	cat
	kus	[kusː]	where	kuss	[kusʲː]	shush
	last	[lɑsːt]	child-SG.PART	last	[lɑsʲːt]	cargo
/t/	kott	[kotː]	worn out shoe	kott	[kotʲː]	bag
	nutt	[nutː]	cry	nutt	[nutʲː]	smarts
	rutt	[rutː]	hurry	Rutt	[rutʲː]	(name)
	pats	[pɑtːs]	slap	pats	[pɑtʲːs]	braid

a small **lamb** was born, not a calf', *Hobustele ehitati uus **tall*** [tɑlʲː], *mis pakub neile peavarju kogu aasta* 'A **stable** was built for the horses, which provides shelter for them throughout the year'.

Twenty-eight participants were recorded in Tartu using an AKG 414 microphone and Focusrite Scarlett 18i20 sound card, and 15 participants were recorded in Tallinn using an Audiotechnica ATM33a microphone and Sound Devices Mixpre 6 or M-Audio Fast Track Pro USB sound card. The recordings were saved as 44.1 kHz/16-bit wave files. In both locations, the recordings were carried out in a soundproof room. All of the speakers spoke wide-spread Standard Estonian. The recording sessions consisted of two trials. In the first trial, the recording was made using an electromagnetic articulograph (EMA; the results were reported in Malmi & Lippus 2019); in the second trial, in order to get more natural speech, the same sentences were read without the five EMA sensors glued on the tongue, and only the acoustic signal was recorded. In this paper, the data from the second trial are analyzed.

2.2 Segmentation of test words

The test words were first automatically segmented using Automatic Speech Recognition-based force alignment software developed at the Tallinn University of Technology (Alumäe, Tilk & Asadullah 2018) and then manually corrected by visually analyzing the corresponding spectrograms. The boundary between the vowel and the following /s/ was set where the periodic oscillation of the vowel became aperiodic. The beginning of the closure phase of /t/ was set where the F2 of the preceding vowel did not show any periodic oscillations anymore. The boundary between a vowel and a following /l/ or /n/ was set by looking at the F2 movement and concentration of energy. For /n/, we drew a boundary where there is a drop in F2 because of the anti-formant that forms in the nasal cavity, which dampens the values in the F2 region. For /l/, the movement of F2 and the relative intensity of the vowel and the lateral were compared, and the boundary was set where the intensity was lower after the vowel. The boundary for the beginning of the vowel was also set by keeping these parameters in mind:

periodicity, the movement of F2, and by looking at the intensity of the spectrogram. The final boundary of /l/ and /n/ was set where the periodic oscillations of F2 ended or became aperiodic. For /s/ and /t/, the boundary was set where the aperiodic oscillations faded out or where the presumable aspiration started.

2.3 Analysis of vowels that precede consonants

The acoustic analysis was carried out using the software Praat (Boersma & Weenink 2021). For the analysis of vowels /ɑ o u/ preceding consonants, an Optimized Formant Ceiling method (Escudero et al. 2009) implemented as a Praat script was used to extract F1, F2, and F3 values from 30 equidistant time points spanning the duration of each vowel. A unique formant ceiling was searched for each vowel between 4000 Hz and 8000 Hz at 10 Hz intervals. The ceiling with the smallest variance within each vowel was chosen for the analysis. After the ceiling was chosen, a new analysis with the selected ceiling was carried out. This allowed us to dynamically analyze the movements of formants with minimalized number of measurement errors. We also measured the total duration of each of the vowels. We analyzed 331 vowel tokens before /l/, 171 before /n/, 172 before /s/ and 237 before /t/. The results concerning spectral properties of vowels are presented in Section 3.1, and for temporal properties in Section 3.3.

2.4 Analysis of consonants

For /n/ and /l/ we used the same formant extraction method as for vowels. For /s/ and /t/, the spectral center of gravity (COG), the spectral standard deviation, and the duration were measured using Praat. The COG is the weighted average of the frequency components of the spectrum over the selected time window from which it is measured; the standard deviation of the spectrum is the square root of the central moment of the spectrum and shows how much the frequencies in a spectrum deviate from the COG. For /s/, the measurements were taken with a 40 ms Hamming window (following the method of Jongman, Wayland & Wong 2000) from the beginning and the middle of the consonant. To avoid overlapping of the measurement points, all segments with a duration below 80 ms were excluded. For /t/, the COG and spectral standard deviation were measured with a 10 ms Hamming window from the beginning and the middle of the release burst. As the burst can be very short, all segments with a duration below 15 ms were excluded from the analysis. We followed the method of Forrest et al. 1988, but used a shorter window as the durations were shorter in our data. The exclusion of data points on the basis of duration was motivated by the fact that two measurement points should not overlap more than 50%. The results for the temporal properties of consonants are presented in Section 3.3. Overlapping of the measurement points could not always be avoided, and we disregarded 6 data points for /t/ out of 349 tokens and 10 for /s/ out of 267 tokens. For /n/ we analyzed 171 tokens and for /l/ 647.

In order to prevent short spectral peaks and remove unnecessary outliers, a 200 Hz cepstral smoothing was applied before the analysis of /s/ and /t/ (see Breithaupt, Gerkmann & Martin 2007). The data were manually checked for very high or low values, and measurement errors were corrected. The results for consonants are presented in Section 3.2.

2.5 Statistical analysis

Statistical data analysis and visualization were carried out using R (R Core Team 2021). A Generalized Additive Mixed Model (GAMM) from the *mgvc* (Wood 2017) package was used to analyze the spectral properties of the vowels /ɑ o u/, and the consonant /l/. A GAMM is a regression model that is able to assess non-linearity in time-varying data. It differs from a linear regression model in that it allows one to smooth the data for a better fit, and it applies penalization to overcome overfitting by adding wiggliness to the regression line where

applicable. For the estimation of smoothing parameters, fREML (fast restricted maximum likelihood estimation) was used; for smoothing, we used cubic regression splines. The residuals of the final models were checked for normality, and the auto correlative structure from the residuals was removed as far as possible by adding an autocorrelation parameter. For a more in-depth description of GAMMs, see Wieling (2018), which we followed for this study.

For the analysis of /n/, we found that it is more suitable to use a linear mixed model (LMM) instead of a GAMM because there was a high degree of linearity in the data. GAMMs that were initially used for the analysis of /n/ also showed that the effect of palatalization was weak. To reduce the risk of over-fitting the model, we took mean values from 30 formant points of each consonant in a word and used that as a dependent variable in LMM. For the analysis of /s/, /t/, and /n/, LMM was used from the *lme4* package (Bates et al. 2015). Linear models assume that there is a linear relationship between dependent and independent variables. They allow for both fixed and random effects.

To compare all pairs of non-palatalized vs. palatalized tokens, we used the formant values (F1, F2, F3) or the log-transformed duration as the dependent variables, and palatalization, vocalic context, and speaker gender as independent variables. After that, we carried out a series of post-hoc tests. The estimated values of non-palatalized and palatalized pairs, standard errors, and *p*-values were acquired by changing the intercept level of the model and then re-running it. In this pairwise comparison, the *p*-values were corrected by using the Bonferroni-Holm method with the *p.adjust* function. In post-hoc testing of segmental durations, the *p*-values were corrected with the Benjamini-Hochberg method. When there were no statistically significant interactions between pairs, an optimal model was chosen to describe the data. In all of the models, random intercepts for participants were included; if it improved the model fit, the test words were also used as a random intercept.

2.6   Formant trajectory visualization

For formant trajectory visualization, GAMM plots were made with the *itsadug* package (van Rij et al. 2017); this plots the contours and the 95% confident intervals from model predictions. We determined the time points (TP) that represent the relative locations where the formant trajectories of non-palatalized and palatalized segments differed significantly from each other. TPs were acquired using the *plot_diff* function from the *itsadug* package; this plots the different curves based on GAMM predictions. The function also gives the time points for which the compared contours differ significantly from each other. These points were added to the formant contour plots and to the model estimate tables. For /l/, we looked at the last timepoint where the difference was significant to assess the scope of palatalization in the consonant. If there was no difference, there was no effect of palatalization, the point was not plotted, and it is marked with an X in the tables.

3   Results

3.1   The quality of the vowels preceding the consonants

The following section analyzes the formants of the vowels preceding /l/, /n/, /s/ and /t/ using GAMMs.

3.1.1   Formant values of the vowels preceding /l/

Figure 1 shows the fitted F1, F2, and F3 values of vowels preceding /l/ from the GAMMs. Table 2 shows the outputs of the post-hoc models with multiple comparisons for each combination of vowel and gender.
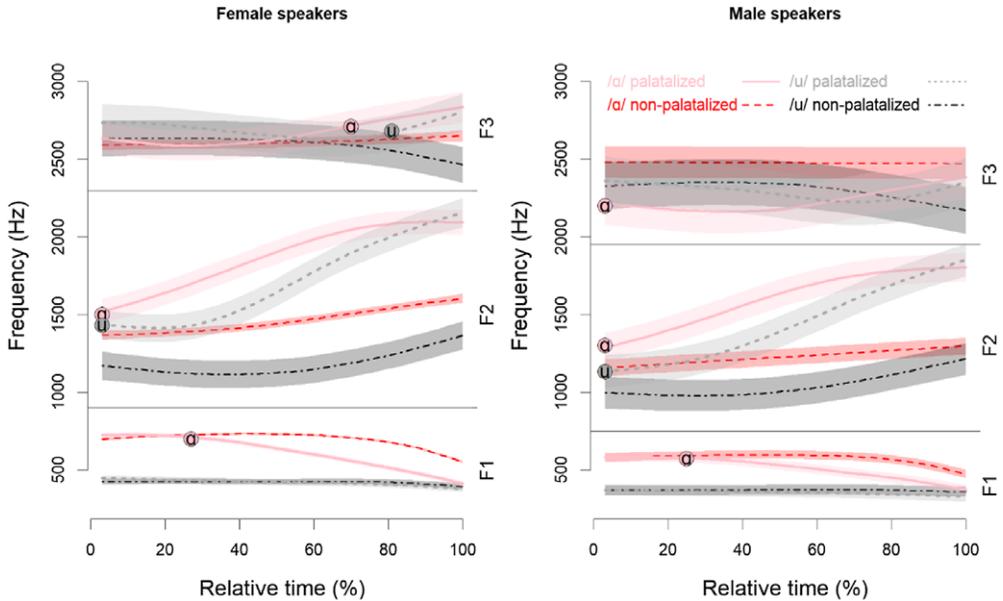
Figure 1 (Colour online) Fitted F1, F2 and F3 values from the GAMMs for the vowels preceding /l/ in test words produced by the female (left panel) and male (right panel) speakers. The light colors represent the vowels in a palatalized context and the dark colors in a non-palatalized context: /ɑ/ – red, /u/ – black. The relative location of the timepoint from where the difference between the palatalized and non-palatalized conditions occur is marked with a dot.

Table 2 The summary of outputs of the post-hoc GAMMs for the vowel formants with pairwise comparisons of palatalization for each vowel and gender. The table presents the estimates of the formant values of the vowels preceding non-palatalized /l/, the standard error (SE), the estimates of formant values of the vowels preceding palatalized /l/ and their standard error; the difference (Diff.) between the non-palatalized and palatalized conditions, the p-value (p) of the significance of the difference and the relative location (%) of the timepoint from where the difference between the palatalized and non-palatalized conditions occur (TP; marked with X if there was no significant TP).

	V1	Gender	Non-palatalized	SE	Palatalized	SE	Diff.	p	TP
F1	/ɑ/	F	702	10.58	603	7.65	−99	< .001	27
	/u/	F	431	11.46	427	9.88	−4	.773	X
	/ɑ/	M	584	11.15	510	7.67	−74	< .001	25
	/u/	M	381	11.99	363	9.91	−18	.064	X
F2	/ɑ/	F	1371	31.10	1811	37.14	440	< .001	3
	/u/	F	1085	37.75	1660	47.95	575	< .001	3
	/ɑ/	M	1129	31.79	1535	37.20	406	< .001	3
	/u/	M	951	38.34	1383	48.05	432	< .001	3
F3	/ɑ/	F	2612	47.57	2685	47.42	73	.357	X
	/u/	F	2583	54.86	2687	61.23	104	.267	X
	/ɑ/	M	2473	49.36	2242	47.50	−231	< .001	3
	/u/	M	2290	56.44	2282	61.35	−8	.885	X

F1 values of /ɑ/ preceding /l/ (Figure 1, Table 2) were approximately 70–100 Hz lower with palatalization for both genders (p < .001). The difference between the palatalized and non-palatalized conditions occurred at 27% of the vowel duration for female speakers and at 25% for male speakers. The F1 of /u/ was not affected by palatalization. F2 values of

vowels were approximately 400–600 Hz higher with palatalization from the beginning of the vowel for both genders (all contexts $p < .001$). F3 values of /ɑ/ were 231 Hz lower from the beginning only for male speakers ($p < .001$). F3 values of /u/ for male speakers and F3 values of /ɑ/ and /u/ for female speakers were not affected by palatalization.

3.1.2 Formant values of the vowels preceding /n/

Figure 2 shows the fitted F1, F2 and F3 values of vowels preceding /n/ from the GAMMs. Table 3 shows the outputs of the post-hoc models with multiple comparisons for each combination of vowel and gender.
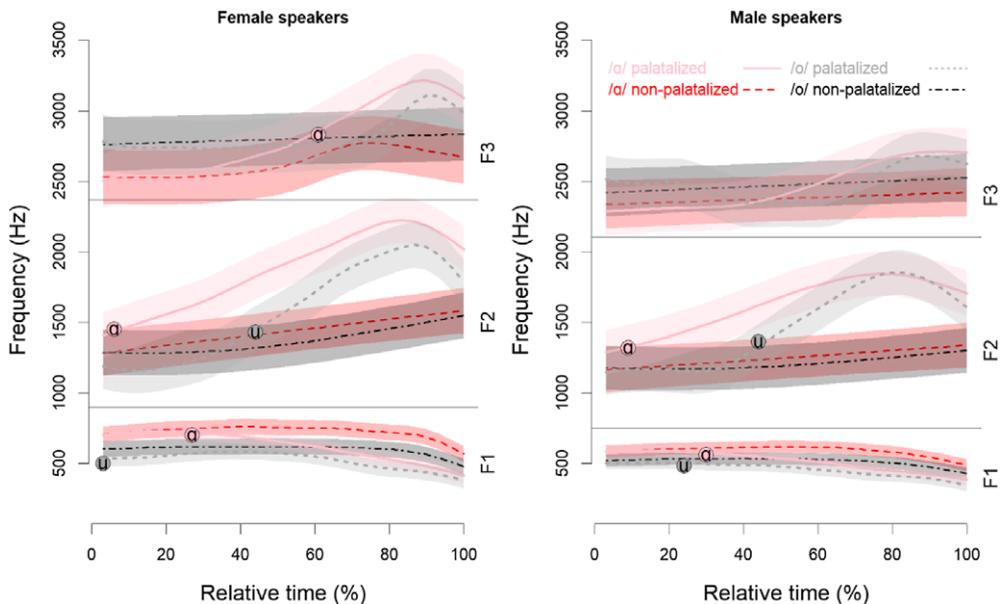

Figure 2 (Colour online) Fitted F1, F2, and F3 values from the GAMMs for the vowels preceding /n/ in test words produced by the female (left panel) and male (right panel) speakers. The light colors represent the vowels in a palatalized context and the dark colors in a non-palatalized context: /ɑ/ – red, /u/ – black. The relative location of the timepoint from where the difference between the palatalized and non-palatalized conditions occur is marked with a dot.

F1 values of /ɑ/ preceding /n/ (Figure 2, Table 3) were approximately 70–90 Hz lower with palatalization for both genders ($p < .001$). F1 of /o/ was 50–70 Hz lower for both genders ($p < .002$). The difference between the palatalized and non-palatalized conditions in /ɑ/ occurred at 27% of the whole duration for female speakers and at 30% for male speakers. The difference in /o/ was at 3% for female speakers and at 24% for male speakers. F2 values of /ɑ/ were 350–410 Hz higher with palatalization for both genders ($p < .001$). F2 values of /o/ were significantly higher for female as well as male speakers (160–210 Hz; $p = .024$ and $p = .002$, respectively). The difference between the palatalized and non-palatalized conditions in /ɑ/ occurred at 6% for females and 9% for male speakers. In the context of /u/, the difference was at 44% for both genders. F3 values of /ɑ/ were 157 Hz higher with palatalization than without for female speakers ($p = .009$). The difference between the palatalized and non-palatalized conditions occurred at 61%. In the other contexts, palatalization did not have an effect on F3 values.

Table 3 The summary of outputs of the post-hoc GAMMs for the vowel formants with pairwise comparisons of palatalization for each vowel and gender. The table presents the estimates of formant values of the vowels preceding non-palatalized /**n**/, the standard error (SE), the estimates of formant values of the vowels preceding palatalized /**n**/ and their standard error; the difference (Diff.) between the non-palatalized and palatalized conditions, the *p*-value (*p*) of the significance of the difference and the relative location (%) of the timepoint from where the difference between the palatalized and non-palatalized conditions occur (TP; marked with X if there was no significant TP).

	V1	Gender	Non-palatalized	SE	Palatalized	SE	Diff.	*p*	TP
F1	/ɑ/	F	732	16.44	639	15.27	−93	< .001	27
	/o/	F	606	16.44	533	15.27	−73	< .001	3
	/ɑ/	M	597	17.63	529	16.20	−68	< .001	30
	/o/	M	521	17.45	468	16.01	−53	.002	24
F2	/ɑ/	F	1395	51.18	1807	69.24	412	< .001	6
	/o/	F	1328	42.82	1489	57.12	161	.024	44
	/ɑ/	M	1212	47.02	1559	61.23	347	< .001	9
	/o/	M	1178	46.74	1389	61.07	211	.002	44
F3	/ɑ/	F	2496	49.41	2653	51.72	157	.009	61
	/o/	F	2677	48.34	2691	50.71	14	.781	X
	/ɑ/	M	2256	52.73	2319	55.18	63	.229	X
	/o/	M	2350	51.84	2357	54.38	7	.856	X

3.1.3 Formant values of the vowels preceding /s/

Figure 3 shows the fitted F1, F2 and F3 values of vowels preceding /s/ from the GAMMs. Table 4 shows the outputs of the post-hoc models with multiple comparisons for each combination of vowel and gender.
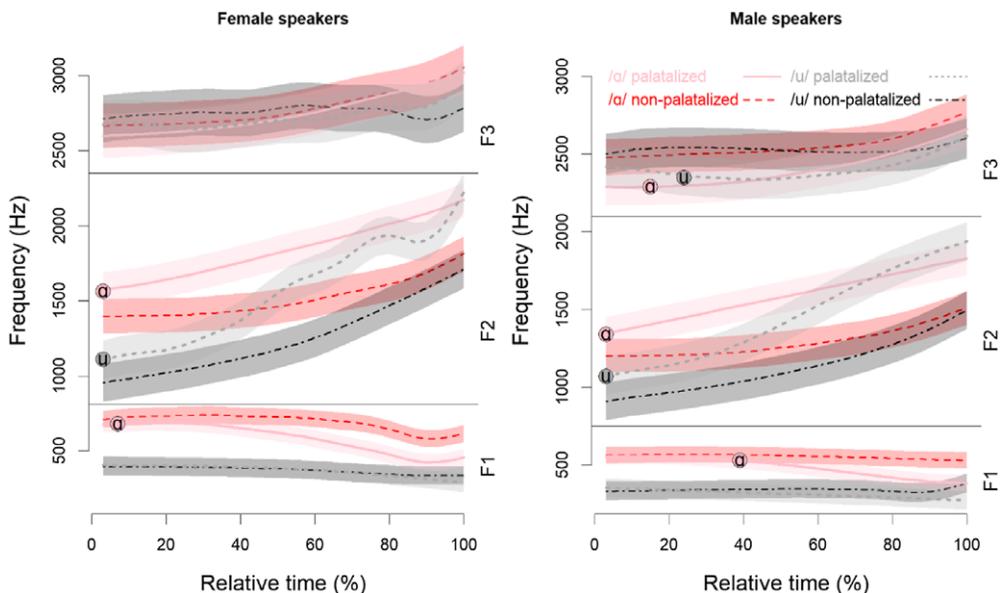

Figure 3 (Colour online) Fitted F1, F2, and F3 values from the GAMMs for the vowels preceding /s/ in test words produced by the female (left panel) and male (right panel) speakers. The light colors represent the vowels in a palatalized context and the dark colors in a non-palatalized context: /ɑ/ – red, /u/ – black. The relative location of the timepoint from where the difference between the palatalized and non-palatalized conditions occur is marked with a dot.

Table 4 The summary of outputs of the post-hoc GAMMs for the vowel formants with pairwise comparisons of palatalization for each vowel and gender. The table presents the estimates of formant values of the vowels preceding non-palatalized /s/, the standard error (SE), the estimates of formant values of the vowels preceding palatalized /sʲ/ and their standard error; the difference (Diff.) between the non-palatalized and palatalized conditions, the *p*-value (*p*) of the significance of the difference and the relative location (%) of the timepoint from where the difference between the palatalized and non-palatalized conditions occur (TP; marked with X if there was no significant TP).

	V1	Gender	Non-palatalized	SE	Palatalized	SE	Diff.	*p*	TP
F1	/ɑ/	F	727	13.67	631	15.46	−96	< .001	7
	/u/	F	404	17.31	404	21.57	0	1	X
	/ɑ/	M	583	14.43	523	16.36	−60	< .001	39
	/u/	M	367	18.63	347	22.86	−20	.437	X
F2	/ɑ/	F	1451	27.86	1757	31.99	306	< .001	3
	/u/	F	1158	35.57	1449	44.98	291	< .001	3
	/ɑ/	M	1232	29.60	1511	33.43	279	< .001	3
	/u/	M	1055	38.15	1357	48.17	302	< .001	3
F3	/ɑ/	F	2663	38.55	2627	30.51	−36	.539	X
	/u/	F	2653	44.02	2613	42.97	−40	1	X
	/ɑ/	M	2444	41.17	2283	32.51	−161	< .001	15
	/u/	M	2431	47.15	2301	45.99	−130	.013	24

F1 values of /ɑ/ preceding /s/ (Figure 3, Table 4) were 60–96 Hz lower with palatalization for both genders (*p* < .001). The difference between the palatalized and non-palatalized conditions in /ɑ/ was at 7% for female speakers and at 39% for male speakers. The F1 value of /u/ was not affected by palatalization. F2 values of vowels were approximately 280–300 Hz higher from the beginning of the vowel with palatalization for both genders (all contexts *p* < .001). Male speakers' F3 values for /ɑ/ were 161 Hz lower (*p* < .001) and those for /u/ were 130 Hz lower (*p* = .013) with palatalization than without. The difference between the palatalized and non-palatalized conditions in /ɑ/ occurred at 15% and in /u/ at 24% of the whole duration. Female speakers' F3 values were not affected by palatalization.

3.1.4   Formant values of the vowels preceding /t/

Figure 4 shows the fitted F1, F2 and F3 values of /t/ from the GAMMs. Table 5 shows the outputs of the post-hoc models with multiple comparisons for each combination of vowel and gender.

F1 values of the /ɑ/ and /o/ preceding /t/ (Figure 4, Table 5) were 34–99 Hz lower with palatalization than without for both genders (all contexts *p* < .001). F1 values of /u/ were 17 Hz higher with palatalization (*p* = .002) for female speakers but not for male speakers. The location of the difference between the palatalized and non-palatalized conditions varied: for /ɑ/ it was at 5% of the vowel for female speakers and at 18% for male speakers; for /o/ at 37% for female speakers and at 32% for male speakers and for /u/ at 38% for female speakers. For male speakers, the difference was not significant. F2 values of the vowels were higher with palatalization for both genders (all contexts *p* < .001). For the vowels /ɑ/, /o/, and /u/, the location of the difference between the two conditions was at 3% of the whole duration. F3 values of vowels preceding /t/ were not affected by palatalization.
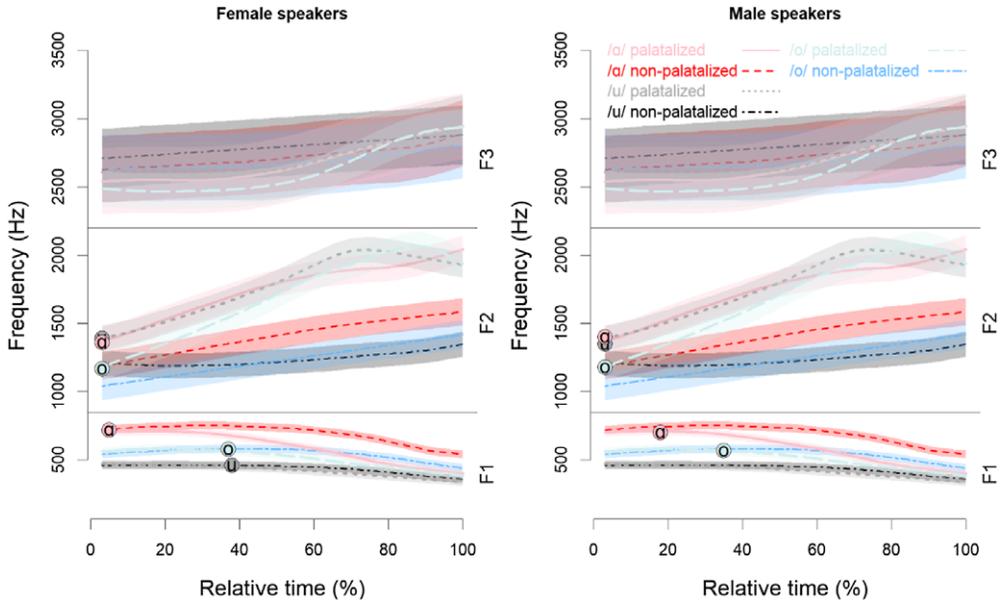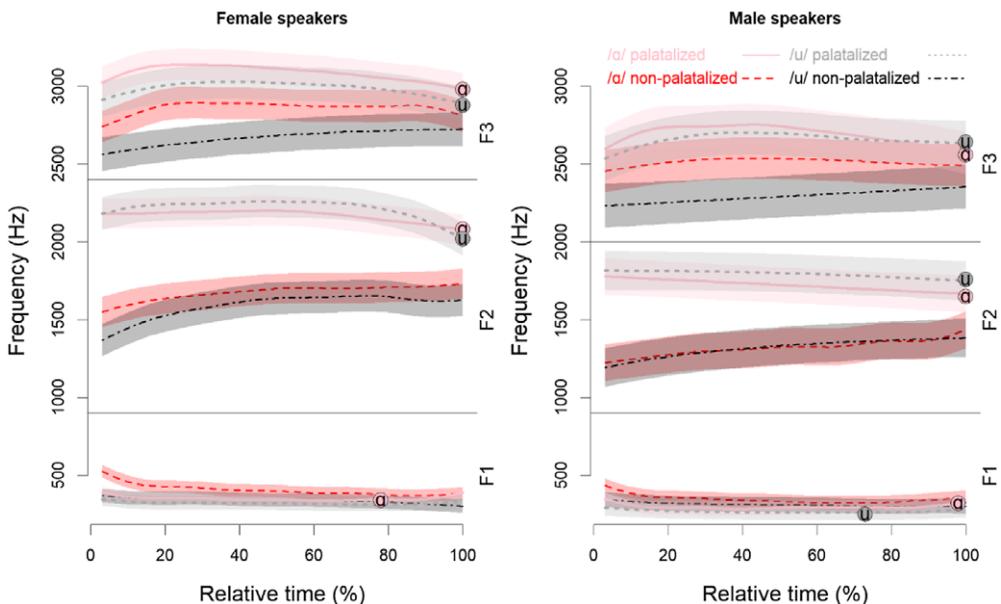
Figure 4 (Colour online) Fitted F1, F2, and F3 values from the GAMMs for the vowels preceding /t/ in test words produced by the female (left panel) and male (right panel) speakers. The light colors represent the vowels in a palatalized context and the dark colors in a non-palatalized context: /ɑ/ – red, /u/ – black and /o/ – blue. The relative location of the timepoint from where the difference between the palatalized and non-palatalized conditions occur is marked with a dot.

Table 5 The summary of outputs of the post-hoc GAMMs for the vowel formants with pairwise comparisons of palatalization for each vowel and gender. The table presents the estimates of formant values of the vowels preceding non-palatalized /t/, the standard error (SE), the estimates of formant values of the vowels preceding palatalized /t/ and their standard error; the difference (Diff.) between the non-palatalized and palatalized conditions, the *p*-value (*p*) of the significance of the difference and the relative location (%) of the timepoint from where the difference between the palatalized and non-palatalized conditions occur (TP; marked with X if there was no significant TP).

	V1	Gender	Non-palatalized	SE	Palatalized	SE	Diff.	*p*	TP
F1	/ɑ/	F	692	9.12	593	6.93	−99	< .001	5
	/u/	F	434	8.46	417	4.99	−17	.002	38
	/o/	F	544	9.42	510	7.33	−34	< .001	37
	/ɑ/	M	587	9.78	504	7.43	−83	< .001	18
	/u/	M	391	9.09	362	5.52	−29	.105	X
	/o/	M	499	10.58	442	8.46	−57	< .001	35
F2	/ɑ/	F	1450	28.04	2141	35.59	344	< .001	3
	/u/	F	1282	22.16	1817	25.81	535	< .001	3
	/o/	F	1275	29.54	1731	36.81	456	< .001	3
	/ɑ/	M	1130	30.11	1572	38.19	442	< .001	3
	/u/	M	1093	23.80	1282	32.52	189	< .001	3
	/o/	M	1046	35.23	1540	42.36	494	< .001	3
F3	/ɑ/	F	2704	82.04	2657	108.68	−47	.595	X
	/u/	F	2779	61.30	2689	76.83	−90	.106	X
	/o/	F	2671	82.81	2594	109.27	−77	.464	X
	/ɑ/	M	2475	83.11	2293	109.33	−182	.941	X
	/u/	M	2498	62.72	2366	77.73	−132	.793	X
	/o/	M	2410	85.20	2382	110.94	−28	.811	X

3.2 Quality of the consonants

The following section analyzes the spectral properties of consonants. The formant values of /l/ were analyzed with GAMMs. For /n/, the GAMMs did not show significant interactions in post-hoc testing, and the data had non-linearity; thus, the mean of 30 formant points was taken and analyzed with a more suitable Linear Mixed Model. COG and the standard deviation of the spectrum of /s/ and /t/ were also analyzed with a Linear Mixed Model; the main effect of palatalization by vocalic context and gender is visualized on the boxplots.

3.2.1 Formant values of /l/

Figure 5 shows the fitted F1, F2, and F3 values of /l/ from the GAMMs. Table 6 shows the outputs of the post-hoc models with multiple comparisons for each combination of vowel and gender.

F1 values of /l/ (Figure 5, Table 6) were 52–63 Hz lower with palatalization than without in the context of /ɑ/ for both genders ($p < .001$). The time point at which non-palatalized and palatalized contours were no longer different occurred at 78% for female speakers and at 98% for male speakers. /l/ in the context of /u/ was 45 Hz lower with palatalization ($p = .05$) with a time point at 73% for male speakers. For female speakers, the difference was not significant. F2 values of /l/ were approximately 420–650 Hz higher with palatalization for both genders and vocalic contexts, and the values stayed higher until the end of /l/ (all $p < .001$). F3 values of /l/ were 181–365 Hz higher with palatalization for

Figure 5 (Colour online) Fitted F1, F2, and F3 values from the GAMMs for /l/ in test words produced by the female (left panel) and male (right panel) speakers. The light colors represent /l/ in a palatalized context and the dark colors in a non-palatalized context: /ɑ/ – red, /u/ – black. The dot marks the relative location of the last time point at which a difference between the palatalized and non-palatalized conditions exists.

Table 6 The summary of outputs of the post-hoc GAMMs for the formants of /l/ with pairwise comparisons of palatalization for each vowel and gender. The table presents the estimates of formant values of non-palatalized /l/, the standard error (SE), the estimates of formant values of palatalized /l/ and their standard error; the difference (Diff.) between the non-palatalized and palatalized conditions, the p-value (p) of the significance of the difference and the relative location (%) of the timepoint from where the difference between the palatalized and non-palatalized conditions occur (TP; marked with X if there was no significant TP).

	V1	Gender	Non-palatalized	SE	Palatalized	SE	Diff.	p	TP
F1	/ɑ/	F	446	11.00	383	12.42	−63	< .001	78
	/u/	F	373	12.66	354	15.97	−19	.675	X
	/ɑ/	M	383	13.35	331	13.35	−52	< .001	98
	/u/	M	345	13.67	300	17.41	−45	.050	73
F2	/ɑ/	F	1626	30.06	2134	27.47	508	< .001	100
	/u/	F	1531	33.15	2178	35.45	647	< .001	100
	/ɑ/	M	1269	32.09	1691	14.54	422	< .001	100
	/u/	M	1267	35.57	1759	38.14	492	< .001	100
F3	/ɑ/	F	2801	36.28	3031	28.12	230	< .001	100
	/u/	F	2604	38.82	2928	36.02	324	< .001	100
	/ɑ/	M	2454	38.71	2635	30.18	181	.005	100
	/u/	M	2230	41.27	2595	38.84	365	< .001	100

both genders in the context of /ɑ/ for female speakers (p < .001) and male speakers (p = .005) and in the context of /u/ (both genders p < .001) and stayed higher until the end of /l/.

3.2.2 Formant values of /n/

Figure 6 shows the formant values of /n/. We used the mean from 30 formant points for this analysis. The outputs of the Linear Mixed Model are reported in the text. The base values in the models were the non-palatalized productions of female speakers' /n/ in the context of /ɑ/.

F1 is presented in the lower panel of figure 6. There was no significant effect of palatalization on the F1 value of /n/. Compared to female speakers ($\beta = 342$ Hz), the value was lower for male speakers ($\beta = -41$ Hz, $t(41) = -3.02$, $p = .004$).

F2 is presented in the middle panel of figure 6. Female speakers' F2 of /n/ (base $\beta = 1696$ Hz) was lower with palatalization ($\beta = -144$ Hz, $t(127) = -2.72$, $p = .007$). In the case of male speakers, F2 was lower compared to female speakers ($\beta = -321$ Hz, $t(62) = -3.65$, $p < .001$). In the interaction with palatalization, the F2 values were higher ($\beta = 250$ Hz, $t(127) = -3.21$, $p = .002$).

F3 of /n/, presented in the upper panel of figure 6, had a significant effect of gender (female base $\beta = 2780$ Hz) and was lower for male speakers ($\beta = -360$ Hz, $t(41) = -4.67$, $p < .001$) but there was no significant effect of palatalization.
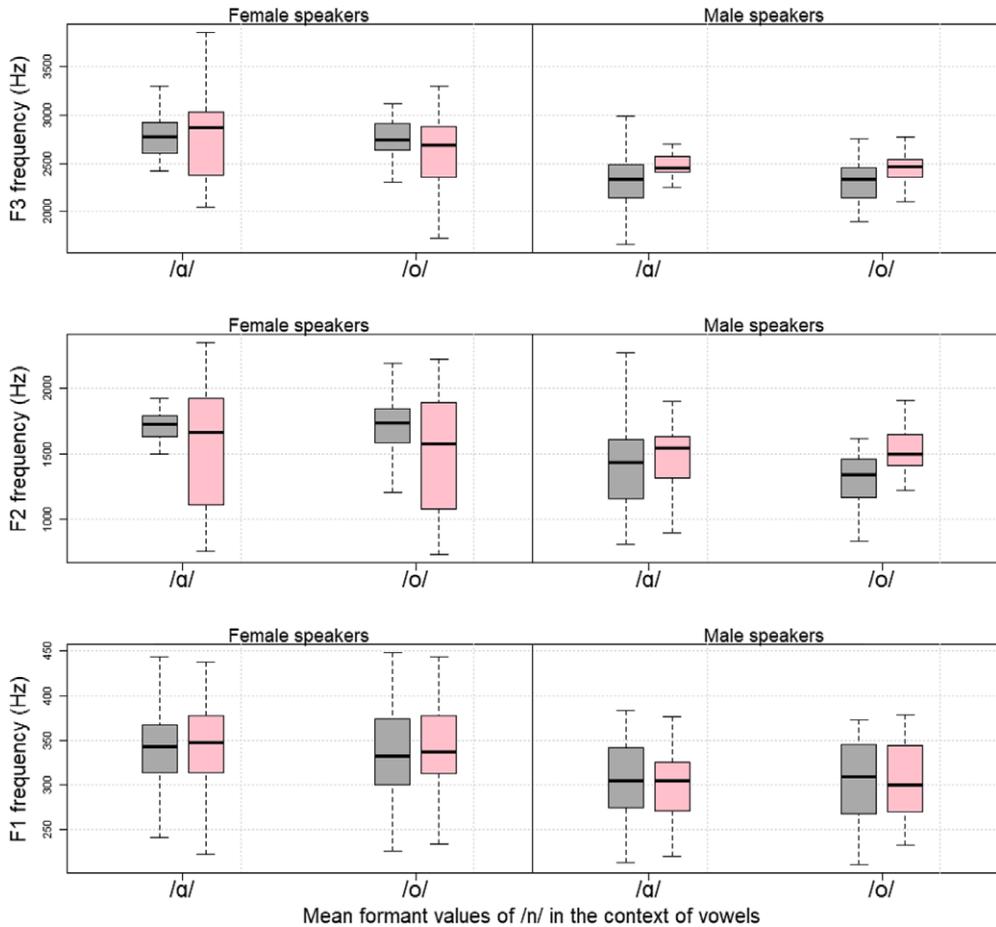
Figure 6 (Colour online) The mean formant values of non-palatalized (gray) and palatalized (pink) /n/ in different vocalic contexts. The top panel is for F3 mean values, the middle one for F2 mean values, and the bottom one for F1 mean values from 30 measurement points of /n/.

3.2.3   Spectral moments of /s/

Figure 7 shows the spectral center of gravity (COG) and the standard deviation of the spectrum measured at the beginning and at the middle of /s/. The base values in the models were the values at the beginning of the non-palatalized /s/ in the context of /ɑ/ of the female speakers.

The COG (presented at the top of Figure 7) of the spectrum of /s/ had a significant main effect of vowel, palatalization, gender and the position of the measurement point. There was also a significant interaction between measurement point and palatalization. Compared to female speakers' non-palatalized /s/ in the context of /ɑ/ (base $\beta = 4402$ Hz), the COG was lower with /u/ ($\beta = -339$ Hz, $t(465) = -4.16$, $p < .001$) and lower with palatalization ($\beta = -550$ Hz, $t(465) = -4.67$, $p < .001$). The COG was lower for male speakers ($\beta = -673$ Hz, $t(41) = -2.99$, $p = .005$) than for female speakers. The COG was higher at the midpoint ($\beta = 946$ Hz, $t(465) = 8.71$, $p < .001$) than at the beginning of /s/. COG values of non-palatalized /s/ were higher at the midpoint with palatalization ($\beta = 593$ Hz, $t(465) = 3.86$, $p < .001$).
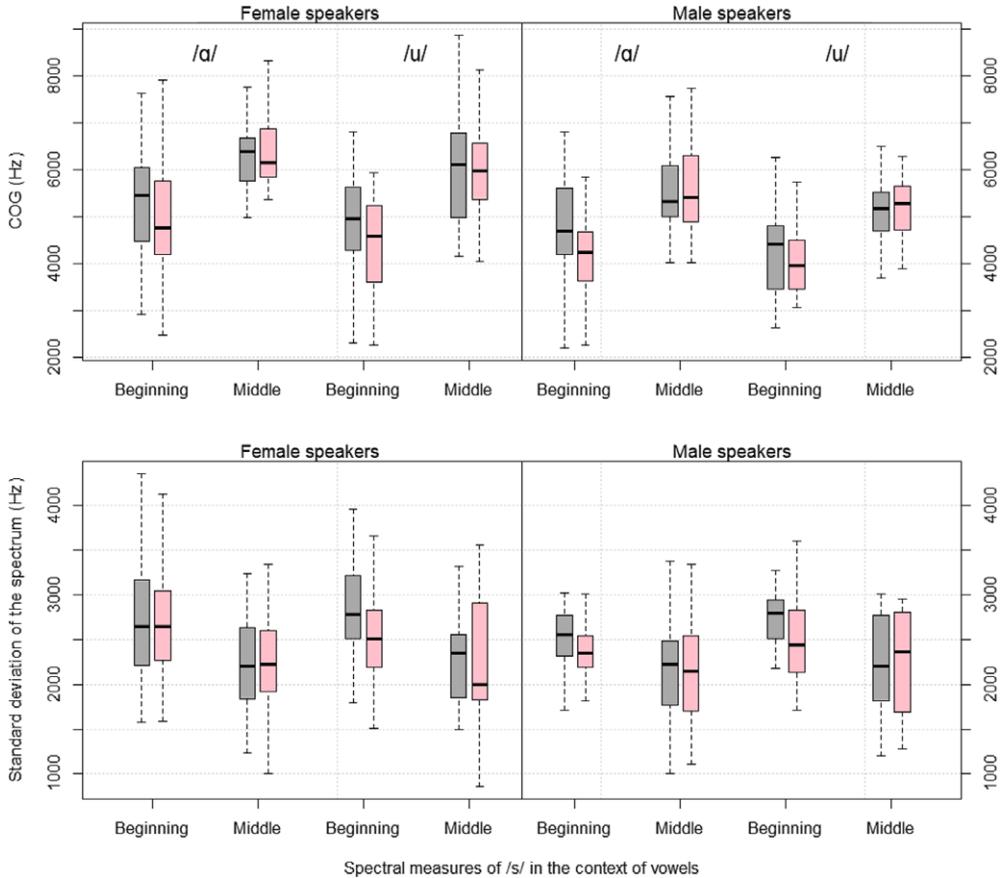
Figure 7 (Colour online) The top panel is COG, and the bottom panel is the standard deviation of the spectrum from the beginning and the middle of /s/ in different vocalic contexts. Gray boxes represent non-palatalized productions, and pink boxes palatalized productions.

The standard deviation of the spectrum (presented at the bottom of figure 7) was not affected by palatalization. Compared to the beginning of /s/ (base $\beta = 2560$ Hz), standard deviation was smaller at the midpoint ($\beta = -334$ Hz, $t(297) = -7.06$, $p < .001$).

3.2.4 Spectral moments of the burst of /t/

Figure 8 shows the spectral center of gravity and the standard deviation of the spectrum measured at the beginning and at the middle of the burst of /t/. The base values in the models were the value at the beginning of the non-palatalized /t/ in the context of /ɑ/ of the female speakers.

The COG of the spectrum of the burst of /t/ is presented in the top panel of Figure 8. There were significant main effects of the vowel and the measurement points and a significant interaction between measurement points and palatalization. Compared to /t/ in the context of /ɑ/ (base $\beta = 5923$ Hz), COG was lower with palatalization ($\beta = -587$ Hz, $t(405) = -2.01$, $p < .045$). The COG was also lower in the context of /o/ ($\beta = -2709$ Hz, $t(402) = -6.96$, $p < .001$) and /u/ ($\beta = -2722$ Hz, $t(403) = -8.04$, $p < .001$). COG was higher at the midpoint ($\beta = 1213$ Hz, $t(405) = 4.37$, $p < .001$) than at the beginning of the burst. Compared to the non-palatalized /t/ in the context of /ɑ/, /t/ in the context of /u/ had a higher COG with palatalization ($\beta = 923$ Hz, $t(407) = 2.31$, $p = .010$). The COG was lower at the midpoint
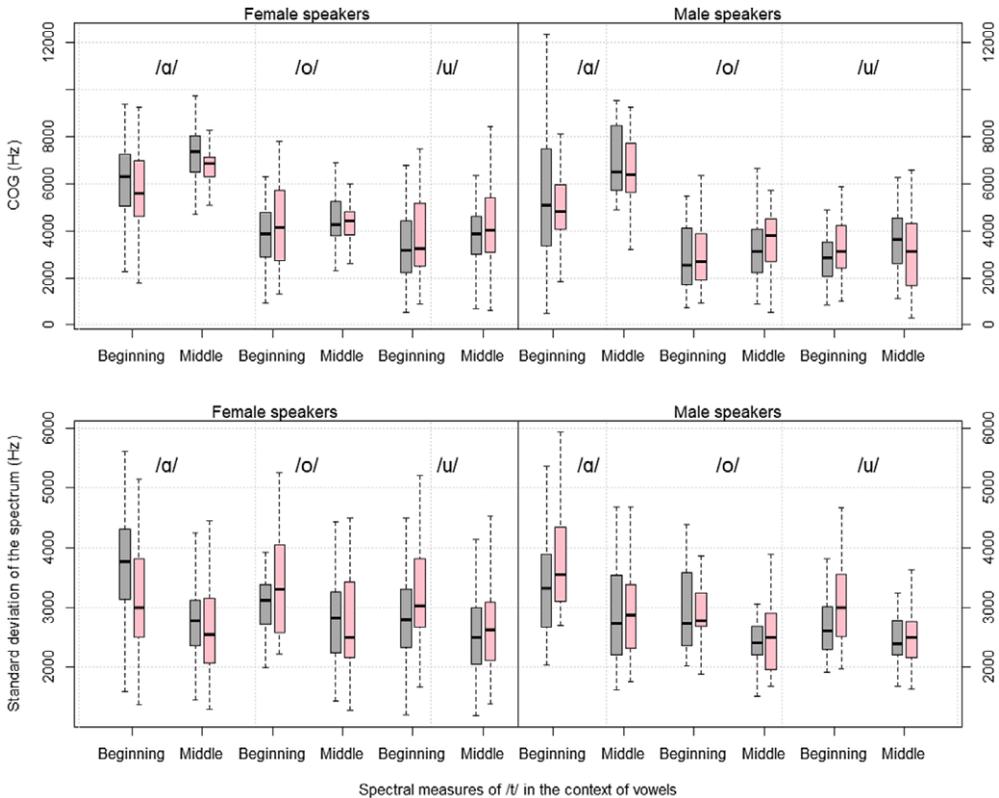
Figure 8 (Colour online) The top panel is the COG, and the bottom panel is the standard deviation of the spectrum from the beginning and middle of the burst of /t/ in different vocalic contexts. Gray boxes represent non-palatalized productions, and pink boxes palatalized productions.

with the vowels /o/ ($\beta = -967$ Hz, $t(406) = -2.31$, $p = .021$) and /u/ ($\beta = -794$, $t(406) = -2.21$, $p = .027$)

The standard deviation of the spectrum of the burst of /t/ is presented in the bottom panel of Figure 8; it was not affected by palatalization. Compared to the beginning of /t/ (base $\beta = 3529$ Hz), standard deviation was smaller at the midpoint ($\beta = -486$ Hz, $t(408) = -6.92$, $p < .001$). The standard deviation of the spectrum was also smaller with the vowels /o/ ($\beta = -681$ Hz, $t(408) = -6.82$, $p < .001$) and /u/ ($\beta = -557$ Hz, $t(408) = -6.48$, $p < .001$).

4 Segmental duration

The following section analyzes the duration of consonants and the preceding vowels with Linear Mixed Models. The duration of vowels are plotted in Figure 9 and the duration of consonants in Figure 10. The mean values with the estimated difference between non-palatalized and palatalized conditions are presented in Table 7.

Post-hoc testing subsequent to a Linear Mixed Model in which the dependent variable was the duration of the vowels showed that /ɑ/ was 22 ms and /u/ was 18 ms longer (both $p = .001$) when preceding /lʲ/. This difference is also visible in Figure 9 and the corresponding mean values are presented in Table 7. /ɑ/ was 32 ms and /o/ 63 ms (both $p = .001$) longer when preceding palatalized /n/; /ɑ/ was 5 ms longer ($p = .001$) when preceding palatalized
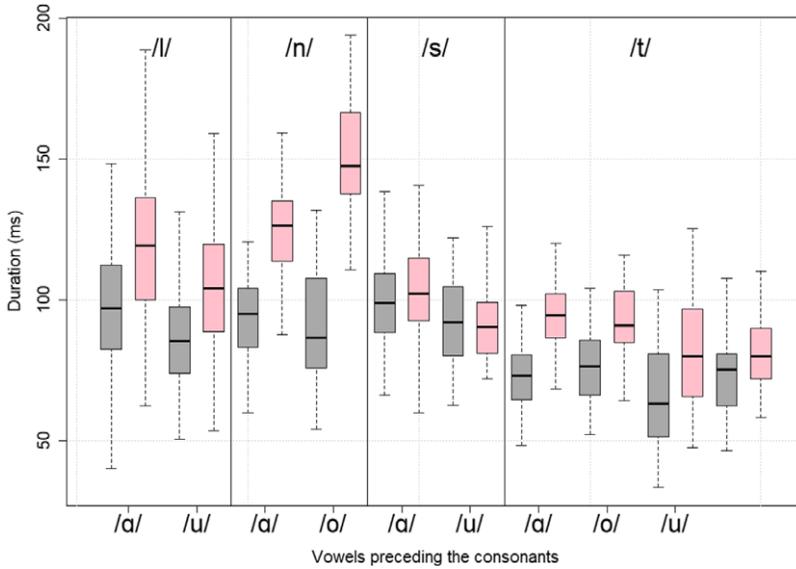
Figure 9 (Colour online) The duration (in milliseconds) of vowels preceding non-palatalized (gray) and palatalized (pink) consonants.
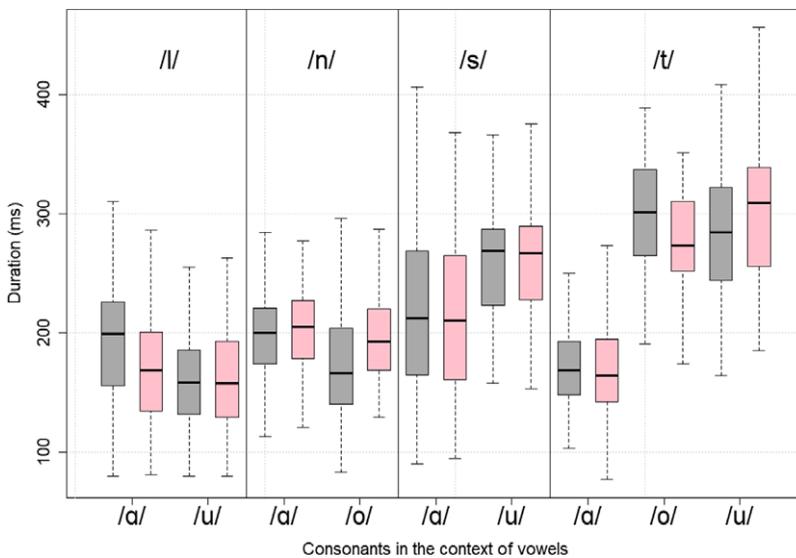

Figure 10 (Colour online) The duration (in milliseconds) of non-palatalized (gray) and palatalized (pink) consonants in the context of vowels.

/s/ but /u/ was not longer in the same context. /ɑ/ was 20 ms, /o/ was 13 ms, /u/ was 16 ms longer (all $p < .001$) when preceding palatalized /t/.

Post-hoc testing subsequent to a Linear Mixed Model where the dependent variable was the duration of the consonant showed that /l/ was 26 ms shorter with palatalization in the context of /ɑ/ ($p < .001$), but not in the context of /u/. This difference is also visible in Figure 10; the corresponding mean values are presented in Table 7. /n/ was 24 ms longer with palatalization in the context of /o/ ($p < .001$), but not in the context of /ɑ/; the duration of /s/ was not affected by palatalization in either vocalic context; the whole duration of /t/

Table 7 The duration (in milliseconds) of vowels and non-palatalized consonants and the durational difference between the non-palatalized and palatalized conditions. The *p*-values in the table are Benjamini–Hochberg corrected.

	C2	/l/		/n/		/s/		/t/		
	V1	/ɑ/	/u/	/ɑ/	/o/	/ɑ/	/u/	/ɑ/	/o/	/u/
Vowels	Non-palatalized	97	86	94	91	98	91	74	78	65
	Palatalized	119	104	126	154	103	91	94	91	81
	Difference	22	18	32	63	5	0	20	13	16
	p	< .001	< .001	< .001	< .001	< .001	1	< .001	< .001	< .001
Consonants	Non-palatalized	195	159	199	173	217	260	172	311	289
	Palatalized	169	160	199	197	216	258	165	276	300
	Difference	−26	1	0	24	−1	−2	−7	−35	11
	p	< .001	1	1	< .001	1	1	1	< .001	1

was 35 ms shorter in the context of /o/ with palatalization ($p < .001$), but not in other vocalic contexts.

5 Discussion

This study investigated the realization of Estonian palatalization by analyzing the spectral and temporal properties of consonants and their preceding vowels.

5.1 The quality of the vowels preceding consonants

In order to capture the fine details of Estonian palatalization, we studied the acoustic cues that describe palatalization in Estonian vowels and quantified the beginning of formant transitions that distinguish non-palatalized and palatalized conditions. We hypothesized that the vowels preceding palatalized consonants would have higher F2, lower F1, and higher F3 values and, because speakers coarticulate segments in speech, it might be difficult to find a clear transition following a stable target segment. We also hypothesized that the formants are continuously moving throughout the whole vowel without a separate [i]-like the transition at the end of the vowel.

As expected, the mean F2 values of vowels in this study were always higher with palatalization, and F1 values tended to be lower. Contrary to the hypothesis, the palatalization effect on F3 values was weak and did not show any systematic variation. As the tongue moves up to the palate, the F2 values rise as the front cavity of the mouth becomes smaller and F1 values lower as the pharyngeal cavity becomes longer and wider as a consequence (Fuchs, Winkler & Perrier 2008, Lee, Shaiman & Weismer 2016). In that, secondary palatalization is realized by the same pattern in Estonian as in other languages.

Previous research on palatalization has focused on describing palatalization through the properties of the transition from the preceding vowel to the consonant, often measuring the quality from a single or a few timepoints (Lehiste 1965, Liiv 1965a, b, Vihman 1967, Derkach et al. 1970, Purcell 1979, Howie 2001, Kochetov 2002, Ćavar 2004, Kim 2012, Ní Chiosáin & Padgett 2012, Teras & Pajusalu 2014, Põld 2016, Piits & Kalvik 2019). We wanted to get a more dynamic view of the movements of the formants and to objectively quantify the time point where the formant values in non-palatalized and palatalized consonants and the vowels preceding them were significantly different. This allowed us to assess how far the scope of palatalization extends into the preceding vowel.

We confirmed our hypothesis that it is difficult to define a stable vocalic segment and a following [i]-like transition. Our results in the current study showed similar results as in

our articulatory study of the same data (Malmi & Lippus 2019), namely that the movement from vowel to consonant was mostly continuous. The F2 values of the vowels that preceded palatalized consonants were almost always higher from the beginning of the vowel compared to those vowels preceding non-palatalized consonants. We attribute this finding to a natural coarticulatory pattern where adjacent segments affect the place of articulation of following segments (Lindblom 1963a, b; Öhman 1966; Amerman et al. 1970; Hammarberg 1976; Lisker 1978; Farnetani & Recasens 2013; Malmi 2014, 2016). At least with short vowels that precede palatalized consonants, we found that the speakers adjust for the movement early on and under-articulate these vowels. The F1 values of vowels were lower with palatalization, but the relative location of the turning point did not show a clear pattern. It varied across consonant contexts and genders in an unclear pattern.

5.2   The quality of consonants

We wanted to observe the possible acoustic cues that describe palatalization in Estonian consonants and to quantify the scope of palatalization. We hypothesized that F2 and F3 values of /l/ are higher and F1 is lower with palatalization. In addition, we expected F2, and F3 values of /n/ would be lower with palatalization, and the COG of /s/ would be lower, and the burst of /t/ would be higher. Additionally, based on a previous articulatory study on the same material (Malmi & Lippus 2019), we hypothesized that the effect of palatalization reaches beyond just the beginning of the consonant as had previously been stated.

As expected, F2 and F3 values of /l/ were higher with palatalization, and F1 values were lower. F2 and F3 values were higher throughout /l/. This also correlates with the findings from our articulatory study, where we found that the place of articulation of /l/ was higher and more fronted with palatalization up until the end of /l/. Higher F2 and F3 values were also shown in previous acoustic studies (Lehiste 1965, Vihman 1967, Remmel & Eek 1971). The rise in F2 shows that the tongue is more fronted and has moved towards the palate. The rise in F3 suggests that the lips might be less protruded or rounded with palatalization.

The F2 values of /n/ were lower with palatalization, while F3 and F1 were not affected. We know from our articulatory study (Malmi & Lippus 2019) that the place of articulation of /n/ is definitely higher and fronted with palatalization, but as it has been noted by other researchers as well (Lehiste 1965, Iskarous & Kavitskaya 2018, Nance & Kirkham 2020), this rise of the tongue body has a weak effect on the F2 of /n/. The lowering of F2 values in our study comes from the anti-formant that forms in the oral cavity and from the dampening effect of the nasal cavity the frequencies in the F2 region (Eek 1972, Tabain et al. 2016). Interestingly, the F2 values for /n^j/ of female speakers varied and had a wider dispersion than male speakers. This is something that we cannot find an explanation for in the context of our study.

The analysis of spectral moments of /s/ confirmed our hypothesis and showed that there was a significant main effect of palatalization where the COG of /s^j/ was lower compared to a non-palatalized context. This is in line with previous studies (Padgett & Żygis 2003, Hamann & Heriberto 2007, Kochetov & Radisic 2009, Malmi 2019). The articulatory data showed that the tongue is higher and fronted with palatalization, and it would be expected that the COG would be higher with palatalization as well. Palatographic studies in Estonian (Ariste 1943, Eek 1973) have shown that with palatalization, the passage through which air can flow over the tongue is narrower, and this might be the reason why the mean spectral energy is lower. Contrary to the preliminary findings of this study (Malmi 2019), we found that the COG was higher at the midpoint of /s/. It might be that the consonant constriction is starting to be released at the midpoint, and the air can thus flow more freely. We also measured the standard deviation from COG from the beginning and the middle of /s/ and whether palatalization had any effect on this. We found that it did not. The only difference was that the standard deviation was smaller in the midpoint than at the beginning of /s/. This suggests that there is more movement at the beginning of /s/ as less in the middle.

The analysis of spectral moments of /t/ showed that there was a lot of variability in the production. Although the place of articulation of /t/ was higher throughout the consonant with palatalization in our articulatory study, the COG was only affected at the beginning of the burst; here, the values were lower. When we looked at the standard deviation of the spectrum of /t/ in the burst phase, we found that it was smaller with palatalization at the midpoint. Palatalization had no effect in the midpoint. A study on Russian (Kochetov 2002) states that the COG is higher in the burst of /t/. It is probably this way because of the high degree of aspiration that comes with it in Russian; however, aspiration is not a feature of palatalization in Estonian.

Our acoustic data suggest that the effect of palatalization on the quality of /n/, /s/, /t/ is weak, and it only affects the quality of the first part of the consonant, as stated above. This means that the most important and persistent cue for palatalization in Estonian is the change in the quality of the preceding vowel. Interestingly, the formant values of /l/ were affected throughout. This might be because of the high degree of sonority, which makes this lateral consonant highly vowel-like.

As consonants in this study varied in regard to the manner of articulation, we struggled to find a unified spectral characteristic that would describe palatalization in consonants in the same way that the rise in F2 describes vowels. The COG values reported in the preliminary paper (Malmi 2019) showed that the values of /l/ were higher and those of /n/ were lower. This follows the same pattern as the F2 values of /l/ and /n/ in this study. It could be concluded that although it is not common to analyze the quality of /l/ and /n/ through the COG, it might be a suitable unified spectral characteristic that describes palatalization in Estonian consonants.

5.3   The duration of vowels and consonants

Thirdly, we wanted to know whether the duration of the vowel is lengthened in the palatalization context and if palatalized consonants are longer than the corresponding non-palatalized consonants. In alignment with what previous research has shown, we hypothesized that the duration of the vowel that precedes palatalized consonants is lengthened. We expected that the duration of /s l n/ and the burst of /t/ are also longer.

The duration of the vowels were, as hypothesized, longer with palatalization. This longer duration has been explained (Lehiste 1965, Liiv 1965a, Kochetov 2006, Ordin 2011, Teras & Pajusalu 2014, Stoll et al. 2015, Piits & Kalvik 2019) by the tongue raising towards the hard palate, which brings about the [i]-like the transition from the vowel into the consonant that accompanies secondary palatalization.

The duration of the consonants varied and did not show a systematic pattern. Both /l/ in the context of /ɑ/, and /t/ in the context of /o/ were shorter, while /n/ in the context of /o/ was longer with palatalization. The duration of /s/ and /t/ in the context of /ɑ/ is noticeably shorter than in other vocalic contexts. This is probably due to the fact that in the context of /ɑ/, the consonant was followed by another consonant, while in other vocalic contexts, it was not. The results are not fully in line with the results from our articulatory study, where we did not find that there was a strong effect of palatalization on the duration of the vowels and consonants. The reason for this inconsistency may lie in the fact that the sensors attached to the tongue in the EMA experiment could have had an effect on the temporal resolution of the tongue movements, adding variability and preventing the effect of palatalization from reaching a significant level. Researchers have found that there is a lengthening effect of palatalization on the duration of /s/ and on the burst of /t/ in other languages (Bolla 1981, Zsiga 2000, Kavitskaya 2006, Stoll et al. 2015).

5.4   Limitations and future perspectives

The current study is limited in that we could not control for the vocalic context of the consonants because we used a large selection of word pairs for the analysis, and thus these could

not make a homogenous group. As the vowels that were analyzed in this study were short, we wonder if palatalization would affect long vowels from the beginning as well.

We analyzed all palatalized–unpalatalized consonant pairs that are available in Standard Estonian. Although they were all alveolar consonants, we chose to treat them separately because they differ in their manner of articulation. This means that we could not analyze the effect of palatalization on the quality of consonants with the same approach, and so we had to use different methods. Further, some of the consonants that we analyzed were in a word final consonant cluster. It could be argued that the consonant that follows the palatalized consonant affects its quality, but our current study and our articulatory study (Malmi & Lippus 2019) shows that there are similar patterns in the realization of palatalization regardless of the following context.

If we compare the mean durations in the current study and in our articulatory study (Malmi & Lippus 2019) that is based on the same data, we find that the durations seemed to be shorter with palatalization in the articulatory study. Although it might have been that the speech tempo was faster in the articulatory study, our results suggest that there are differences in how temporal properties of palatalization are realized in those two test conditions. Analyzing the effect of the articulograph on the quality of speech would be an interesting topic to look into as well.

6   Conclusions

In this study, the spectral and temporal properties of non-palatalized /l t n s/ and palatalized /lʲ tʲ nʲ sʲ/ and the preceding vowels were analyzed. For vowels and the consonants /l/ and /n/, the dynamics of F1, F2, and F3 values were analyzed; for /s/ and /t/, the center of gravity (COG) and the standard deviation from the COG were measured.

We applied a dynamic formant analysis on the vowels that preceded consonants and wanted to assess the scope of palatalization in the vowels by using a method based on statistics. The results showed that palatalization increased the F2 values of the vowels such that these values were almost always higher from the beginning of the vowels. F1 values decreased with palatalization, but the effect was weaker, and the scope was not as far-reaching as for F2. We found that speakers initiate the palatalization gesture early on and adjust their articulators in anticipation.

A formant analysis of /l/ showed an increase in the F2 and F3 values with palatalization, and the values continued to be higher throughout the lateral consonant, while F1 values were lower. F2 values of /n/ were lower with palatalization. The COG of /s/ was lower at the beginning of the segment with palatalization. At the midpoint of /s/, the COG values were higher with palatalization. The COG of the burst of /t/ was lower at the beginning with palatalization.

The durations of the vowels preceding palatalized consonants were longer, just as expected. However, the data did not show any systematic variation of duration in consonants with palatalization; this was unexpected.

The data suggest that the most persistent acoustic cues for palatalization of consonants in Estonian are the change in the quality and the lengthened duration of the preceding vowel. The effect of palatalization on the consonant is weaker and seems to only affect it in the first part. This was unexpected because the results from our articulatory study on the same data showed that the tongue stayed higher throughout the consonant. Acoustic and articulatory data are not always comparable and need to be complemented by each other to get a complete picture of how palatalization is realized.

References

Alumäe, Tanel, Ottokar Tilk & Asadullah. 2018. Advanced rich transcription system for Estonian speech. In Kadri Muischnek & Kaili Müürisep (eds.), *Human Language Technologies – the Baltic Perspective: Proceedings of the Eighth International Conference*, 1–8. Amsterdam: IOS Press.

Amerman, James D., Raymond Daniloff & Kenneth L. Moll. 1970. Lip and jaw coarticulation for the phoneme /ʃ/. *Journal of Speech, Language and Hearing Research* 13, 147–161.

Ariste, Paul. 1943. *Katselisfoneetilisi tähelepanekuid: mit einem Referat: Experimentalphonetische Beobachtungen*. Tartu: Acta Universitatis Tartuensis (Dorpatensis).

Asu, Eva Liina & Pire Teras. 2009. Illustrations of the IPA: Estonian. *Journal of the International Phonetic Association* 39(3), 367–372. https://doi.org/10.1017/S002510030999017X.

Bateman, Nicoleta. 2007. *A crosslinguistic investigation of palatalization*. Ph.D. dissertation, University of California San Diego.

Bateman, Nicoleta. 2011. On the typology of palatalization. *Language and Linguistics Compass* 5(8), 588–602. https://doi.org/10.1111/j.1749-818X.2011.00294.x.

Bates, Douglas, Martin Mächler, Ben Bolker & Stewen Walker. 2015. Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software* 67(1), 1–48. https://doi.org/doi:10.18637/jss.v067.i01.

Battisti, Elisa & Ben Hermans. 2020. The structural motivation of palatalization. *Fórum Linguístico* 17(Núm. Esp.), 4596–4611. https://doi.org/10.5007/1984-8412.2020v17nespp4596.

Bhat, Shankara D. N. 1978. A general study of palatalization. In Joseph H. Greenberg, Charles A. Fergusion & Edith A. Moravcsik (eds.), *Universals of human language*, vol. 3: *Word structure*, 47–92. Stanford, CA: Stanford University Press.

Boersma, Paul & David Weenink. 2021. Praat: Doing phonetics by computer. [Computer program]

Bolla, Kalman. 1981. *A conspectus of Russian speech sounds*. Cologne: Bölau Verlag.

Botinis, Antonis, Anthi Chaida & Evgenia Magoula. 2011. Phonology and phonetics of Greek palatalisation. *ExLing 2011: Proceedings of 4th Tutorial and Research Workshop on Experimental Linguistics*, vol. 4, 43–46. https://doi.org/10.36505/exling-2011/04/0010/000179.

Breithaupt, Colin, Timo Gerkmann & Rainer Martin. 2007. Cepstral smoothing of spectral filter gains for speech enhancement without musical noise. *IEEE Signal Processing Letters* 14(12), 1036–1039. https://doi.org/10.1109/LSP.2007.906208.

Catford, J. C. 1977. *Fundamental problems in phonetics*. Edinburgh: Edinburgh University Press.

Ćavar, Małgorzata Ewa. 2004. *Palatalization in Polish: An interaction of articulatory and perceptual factors*. Ph.D. dissertation, University of Potsdam.

Chomsky, Noam & Morris Halle. 1968. *The sound pattern of English*. New York: Harper & Row.

Derkach, Miron, Gunnar Fant & Antonio de Serpa-Leitao. 1970. Phoneme coarticulation in Russian hard and soft VCV-utterances with voiceless fricatives. *STL-QPSR* 11(2–3), 1–7.

Draxler, Christoph & Klaus Jänsch. 2004. Speechrecorder: A universal platform independent multi-channel audio recording software. Presented at the 4th International Conference on Language Resources and Evaluation, Lisbon.

Eek, Arvo. 1972. Acoustical description of the Estonian sonorant types. *Estonian Papers in Phonetics* 1, 9–37.

Eek, Arvo. 1973. Observations in Estonian palatalization: An articulatory study. *Estonian Papers in Phonetics* 2, 9–17.

Escudero, Paola, Paul Boersma, Andréia Schurt Rauber & Ricardo A. H. Bion. 2009. A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *The Journal of the Acoustical Society of America Society of America* 126, 1379–1393. https://doi.org/10.1121/1.3180321.

Fant, Gunnar. 1960. *Acoustic theory of speech production*. The Hague: Mouton.

Farnetani, Edda & Daniel Recasens. 2013. Coarticulation and connected speech processes. In William J. Hardcastle, John Laver & Fiona E. Gibbon (eds.), *The handbook of phonetic sciences*, 316–351. Oxford: Wiley-Blackwell.

Forrest, Karen, Gary Weismer, Paul Milenkovic & Ronald N. Dougall. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *The Journal of Acoustical Society of America* 84(1), 115–123.

Fuchs, Susanne, Ralf Winkler & Pascal Perrier. 2008. Do speakers' vocal tract geometries shape their articulatory vowel space? In Rudolph Sock, Susanne Fuchs & Yves Laprie (eds.), *8th International Seminar on Speech Production*, 333–336. Strasbourg: INRIA.

Hamann, Silke & Avelino Heriberto. 2007. An acoustic study of plain and palatalized sibilants in Ocotepec Mixe. *Proceedings of the 16th International Congress of Phonetic Sciences* (ICPhS XVI), Saarbrücken, 949–952.

Hammarberg, Robert. 1976. The metaphysics of coarticulation. *Journal of Phonetics* 4, 353–363.

Howie, Stephen M. 2001. Formant transitions of Russian palatalized and nonpalatalized syllables. *IULC Working Papers* 1(1), https://scholarworks.iu.edu/journals/index.php/iulcwp/article/view/25725/31470.

Hume, Elizabeth V. 1994. *Front vowels, coronal consonants, and their interaction in nonlinear phonology*. New York: Routledge.

Iskarous, Khalil & Darya Kavitskaya. 2018. Sound vhange and the structure of synchronic variability: Phonetic and phonological factors in Slavic palatalization. *Language* 94(1), 43–83. https://doi.org/10.1353/lan.0.0226.

Jongman, Allard, Ratree Wayland & Serena Wong. 2000. Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America* 108(3), 1252. https://doi.org/10.1121/1.1288413.

Kask, Arnold. 1972. *Eesti keele ajalooline grammatika* [Historical grammar of Estonian]. Tartu: Tartu Ülikool, eesti keele kateeder.

Kavitskaya, Darya. 2006. Perceptual salience and palatalization in Russian. In Louis Goldstein, D. H. Whalen & Catherie T. Best (eds.), *Laboratory Phonology*, vol. 8, 589–610. New York: De Gruyter Mouton.

Kim, Hyunsoon. 2012. Gradual tongue movements in Korean Palatalization as coarticulation: New evidence from stroboscopic cine-MRI and acoustic data. *Journal of Phonetics* 40(1), 67–81. https://doi.org/10.1016/j.wocn.2011.07.004.

Kochetov, Alexei. 2002. *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization*. Toronto: University of Toronto.

Kochetov, Alexei. 2006. Testing licensing by cue: A case of Russian palatalized coronals. *Phonetica* 63(2–3), 113–148. https://doi.org/10.1159/000095305.

Kochetov, Alexei. 2011. Palatalization. In Marc van Oostendorp, Colin J. Ewen, Elizabeth V. Hume & Keren (eds.), *The Blackwell companion to phonology*, vol. III, 1666–1690. Oxford: John Wiley & Sons. https://doi.org/10.1002/9781444 335262.wbctp0071.

Kochetov, Alexei & Milica Radisic. 2009. Latent consonant harmony in Russian: Experimental evidence for agreement by correspondence. In Maria Babyonyshev, Darya Kavitskaya & Jodi Reich (eds.), *Seventeenth Formal Approaches to Slavic Linguistics Meeting*, 111–130. Ann Arbor, MI: Jindřich Toman's Michigan Slavic Publications.

Lee, Jimin, Susan Shaiman & Gary Weismer. 2016. Relationship between tongue positions and formant frequencies in female speakers. *The Journal of the Acoustical Society of America* 139(1), 426. https://doi.org/10.1121/1.4939894.

Lehiste, Ilse. 1965. Palatalization in Estonian: Some acoustic observations. *Estonian poetry and language: Studies in honor of Ants Oras*, 136–162. Stockholm: Vaba Eesti.

Liiv, Georg. 1965a. Preliminary remarks on the acoustic cues for palatalization in Estonian. *Phonetics* 13, 59–64.

Liiv, Georg. 1965b. Some experiments on the effect of vowel–consonant transitions upon the perception of palatalization in Estonian. *Soviet Fenno-Ugric Studies* 1, 33–36.

Lindblom, Björn E. 1963a. *On vowel reduction* (Report no. 29). Stockholm: The Royal.

Lindblom, Björn E. 1963b. Spectrographic study of vowel reduction. *The Journal of the Acoustical Society of America* 35, 1773–1781.

Lisker, Leigh. 1978. Segment duration, voicing and the syllable. In Alan Bell & Joan B. Hopper (eds.), *Syllables and segments*, 133–142. Amsterdam: North-Holland.

Malmi, Anton. 2014. Intervokaalse /l/-i kvaliteet ja kvantiteet [The quality and quantity of the intervocalic /l/]. BA thesis, University of Tartu.

Malmi, Anton. 2016. Intervokaalse /l/-i kvaliteet ja palatograafia [The quality and palatography of the intervocalic /l/]. MA thesis, University of Tartu.

Malmi, Anton. 2019. Spectral characteristics of Estonian palatalization. In Sascha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (ICPhS XIX), Melbourne, 3782–3786. Canberra: Australasian Speech Science and Technology Association.

Malmi, Anton & Pärtel Lippus. 2019. Keele asend eesti palatalisatsioonis [The position of the tongue in Estonian palatalization]. *Eesti ja soome-ugri keeleteaduse ajakiri – Journal of Estonian and Finno-Ugric Linguistics* 10(1), 105–128. https://doi.org/10.12697/jeful.2019.10.1.06. [University of Tartu Press]

Nance, Claire & Sam Kirkham. 2020. The acoustics of three-way lateral and nasal palatalisation contrasts in Scottish Gaelic. *The Journal of the Acoustical Society of America* 147(4), 2858–2872. https://doi.org/10.1121/10.0000998.

Ní Chiosáin, Máire & Jaye Padgett. 2012. An acoustic and perceptual study of Connemara Irish palatalization. *Journal of the International Phonetic Association* 42(2), 171–191. https://doi.org/10.1017/S0025100312000059.

Öhman, Sven Erik Gustav. 1966. Coarticulation in VCV utterances: Spectrographic measurements. *The Journal of the Acoustical Society of America* 39, 151–168. https://doi.org/10.1121/1.1909864.

Ordin, Mikhail. 2010. Palatalization and temporal organisation of CVC clusters in Russian. *Russian linguistics* 34(1), 57–65. https://doi.org/10.1007/s11185-009-9049-4.

Ordin, Mikhail. 2011. Palatalization and intrinsic prosodic vowel features in Russian. *Language and Speech* 54(4), 547–568. https://doi.org/10.1177/0023830911404962.

Padgett, Jaye & Marzena Żygis. 2003. The evolution of sibilants in Polish and Russian. *ZAS Papers in Linguistics* 32, 155–174.

Piits, Liisi & Mari-Liis Kalvik. 2019. Palatalisatsioon ühesilbilistes i-tüvelistes pika vokaaliga sõnades [Palatalization in monosyllabic i-stemmed words with a long vowel]. *Keel ja kirjandus* 7, 513–533.

Põld, Enel. 2016. *Pika vokaali järgse konsonandi hääldus i-tüvelistes sõnades eesti murretes* [Pronunciation of a consonant after a long vowel in i-stemmed words]. Tartu Ülikool.

Purcell, Edward T. 1979. Formant frequency patterns in Russian VCV utterances. *The Journal of the Acoustical Society of America* 66(6), 1691–702.

R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/ (29 November 2018).

Remmel, Mart & Arvo Eek. 1971. The laterals: An acoustical study. *9th Acoustics Conference: Psychological Acoustics and Psychoacoustics*, Bratislava, 138–141.

Shupljakov, V., G. Fant & A. De Serpa-Leitao. 1968. Acoustical features of hard and soft Russian consonants in connected speech: A spectrographic study. *STL-QPSR* 9(4), 1–6.

Stoll, Taja, Jonathan Harrington & Phil Hoole. 2015. Intergestural organisation and CV-overlap in palatalised liquids in Russian. *18th International Congress of Phonetic Sciences* (ICPhS XVIII), Glasgow, Scotland.

Tabain, Marija, Andrew Butcher, Gavan Breen & Richard Beare. 2016. An acoustic study of nasal consonants in three Central Australian languages. *The Journal of the Acoustical Society of America* 139(2), 890–903. https://doi.org/10.1121/1.4941659.

Teras, Pire & Karl Pajusalu. 2014. Palatalization and prepalatalization in Estonian spontaneous speech. *Keel ja kirjandus* 4, 257–269.

van Rij, Jacolien, Martijn Wieling, Harald Baayen & Hedderik van Rij. 2017. *itsadug*: Interpreting time series and autocorrelated data Using GAMMs.

Vihman, Marylin M. 1967. *Palatalization in Russian and Estonian* (Project on Linguistic Analysis Reports V1–V32). Berkley, CA.

Wieling, Martijn. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86–116.

Wood, Simon N. 2017. *Generalized additive models: An introduction with R*, 2nd edn. New York: Chapman and Hall/CRC.

Xu, Yi. 2020. Syllable is a synchronization mechanism that makes human speech possible. *PsyArXiv*. https://doi.org/10.31234/osf.io/9v4hr.

Zsiga, Elizabeth C. 2000. Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics* 28(1), 69–102. https://doi.org/10.1006/jpho.2000.0109.